Evidential Generative Adversarial Networks for handling imbalanced learning

Fares Grina[1,2], Zied Elouedi[1], and Eric Lefevre[2]

[1] Université de Tunis, Institut Supérieur de Gestion de Tunis, LARODEC, Tunisia
grina.fares2@gmail.com, zied.elouedi@gmx.fr
[2] Univ. Artois, UR 3926, Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A), F-62400 Béthune, France
eric.lefevre@univ-artois.fr

Abstract. The predictive performance of machine learning models tends to deteriorate in the presence of class imbalance. Multiple strategies have been proposed to address this issue. A popular strategy consists of oversampling the minority class. Classic approaches such as SMOTE utilize techniques like nearest neighbor search and linear interpolation, which can pose difficulties when dealing with datasets that have a large number of dimensions and intricate data distributions. As a way to create synthetic examples in the minority class, Generative Adversarial Networks (GANs) have been suggested as an alternative technique due to their ability to simulate complex data distributions. However, most GAN-based oversampling methods tend to ignore data uncertainty. In this paper, we propose a novel GAN-based oversampling method using evidence theory. An auxiliary evidential classifier is incorporated in the GAN architecture in order to guide the training process of the generative model. The objective is to push GAN to generate minority objects at the borderline of the minority class, near difficult-to-classify objects. Through extensive analysis, we demonstrate that the proposed approach provides better performance, compared to other popular methods.

Keywords: Imbalanced classification · Generative models · Oversampling · Dempster-Shafer theory

1 Introduction

Unequal amount of data in different classes can cause many issues with classification performance. Due to this imbalance, conventional classifiers tend to focus on the majority class and overlook the minority class. However, this latter can often contain important information that needs to be carefully analyzed in real-world scenarios, such as intrusion detection [9], medical diagnosis [22], fraud detection [2], and satellite data analysis [7]. This machine learning problem attracted significant interest [12], investigating the question of how to make learning algorithms acquire unbiased knowledge from imbalanced data. Most models face difficulty distinguishing minority classes and often treat them as "noise" in comparison to majority classes when the training data is heavily biased towards one or a few classes. The problem gets more difficult by the fact that standard measures like accuracy can be deceptive in assessing the model. For instance, if a model simply assigns the majority class to all samples, it may still have a high accuracy score when the class distribution is heavily imbalanced.

Different methodologies have been proposed to address this issue. Mainly, the leading approaches are resampling, cost-sensitive algorithms, and ensemble methods. Resampling generally consists of oversampling the minority class by adding synthetic data, or undersampling the majority class by removing data. Oversampling is one of the most proven methods for handling class imbalance [12]. Other than random oversampling (randomly selecting and replicating minority data), the Synthetic Minority Oversampling Technique (SMOTE) [6] is a classic oversampling choice. The SMOTE technique firstly selects at random a minority object and a nearest neighbor example from the minority class at random. An important limitation of this method is the fact that it only considers the minority class, which means that the relationship between the minority class and the majority class is overlooked. This makes this method inefficient in many scenarios, especially when there are other data difficulties in the dataset, such as high uncertainty (such as class overlapping and noise).

To address this drawback, many SMOTE-based variants have been suggested over the years. BorderlineSMOTE [15] and ADASYN [16] are very similar to SMOTE, but with control over the locations of generated minority examples. More recently, other methods based on SMOTE paired with undersampling tackled class overlapping problem in imbalanced data [14, 19]. However, most SMOTE-based techniques are based on non-parametric models such as the k-nearest neighbors (k-NN) [8], which makes them not very efficient when dealing with high dimensional and complex datasets.

More recently, Generative Adversarial Networks (GANs) [13] have emerged as a type of deep generative model, which goal is to reconstruct the real data distribution and generate a synthetic one. GANs have been used as an oversampling method to generate minority class instances, in order to rebalance the data. Although the majority of GAN studies concentrate on unstructured, continuous data like images and text, most classification datasets in real-world business situations consist of tabular data (numerical and categorical data). Very few proposals have addressed this type of data in GAN-based oversampling literature [11]. Some works use unorthodox strategies to deal with tabular data, such us converting it into two-dimensional in order to be processed by 2d-convolutions [28], which is not always the best solution [11]. Other than types of data, most GAN-based oversampling methods are developed to generate data without taking into account the uncertainty present in the data. The class imbalance issue has been proven to get worse in the presence of ambiguity [14,35]. Thus, it is important for GAN-based oversampling to generate data near the borderline of the minority and majority classes. This aspect holds significant importance as it helps address the challenge of imbalanced datasets by focusing on the regions where the minority class is particularly vulnerable (high uncertainty).

In this paper, we propose an evidential GAN-based oversampling method, that can enhance the robustness of the minority class borders, by generating boundary samples. The theory of Evidence [32] was used to guide the training of our GAN, in order to simulate the distribution of boundary minority objects. The intuition is to introduce highly uncertain data in the minority, for the purpose of empowering the difficult-to-classify objects. This is done by modifying the GAN architecture by adding an auxiliary evidential component to feed uncertainty information to the generator. To incorporate auxiliary knowledge and guide the training of the GAN, two regularization terms are introduced into the loss function. These regularization terms serve the purpose of leveraging additional information to enhance the learning process. A mechanism is also implemented in order to effectively model tabular data with numerical and categorical features.

The remainder of this paper will be divided as follows. Firstly, we provide some background information for GANs and theory of evidence in Section 2. Section 3 presents our proposal, detailing each step. Experimental evaluation and discussion are conducted in Section 4. Our paper ends with a conclusion and an outlook on future work in Section 5.

2 Preliminaries

Before introducing our approach, we firstly present some necessary background information.

2.1 Generative Adversarial Networks (GANs)

GANs [13] are composed of two neural networks that work against each other. One of these networks is the generator G, which maps a low-dimensional latent space to a high-dimensional sample space of x. The second network is the discriminator D, which acts as a binary classifier to distinguish real inputs from fake inputs generated by the generator G. The generator and discriminator are trained in an alternating manner to minimize the following min-max loss:

$$\min_G \max_D L(D,G) = \mathbb{E}_{x\sim p_{real}}\left[log(D(x))\right] + \mathbb{E}_{z\sim p_z}\left[log(1-D(G(z)))\right] \quad (1)$$

where z is the noise input to G, usually following a normal distribution p_z, and x is an example from the real dataset p_{real}. The objective functions of discriminator D and generator G are as follows:

$$L_D = \mathbb{E}_x[log(D(x))] + \mathbb{E}_z[log(1-D(G(z)))] \quad (2)$$

$$L_G = \mathbb{E}_z[log(D(G(z)))] \quad (3)$$

Conditional GAN (cGAN) [24] extended the vanilla version by allowing the conditioning of G and D. For example, one can add a class condition to the input of the generator, to ensure that the generated objects belong to the chosen class. At the same time, the condition helps the discriminator to make more informed predictions.

Although successful, GANs are known to be difficult to train, which is a phenomenon called mode collapse. Indeed, GANs aim to generate a variety of outputs, but if a generator produces a highly plausible output, it may learn to produce only that output. If the discriminator consistently rejects that output, the generator may get stuck producing a small set of similar outputs. Among many approaches addressing this issue, Fisher GAN [26] is a type of GAN that uses the Fisher distance as the metric to measure the distance between the distributions of the generated and real data. It modifies the objective function of the GAN by replacing the discriminator with the score function of the generator, and minimizing the Fisher distance between the generated and real data distributions. This leads to more stable training and reduces the risk of mode collapse, compared to traditional GANs. In this paper, a conditional version of Fisher GAN will be used, with the condition being on the minority class.

2.2 Evidential uncertainty quantification

The theory of evidence [10,32,33], also known as Dempster-Shafer theory (DST) or belief function theory, provides a robust and adaptable framework to represent and merge uncertain information. Let $\Omega = \{w_1, w_2, ..., w_K\}$ be a frame of discernment composed of a finite set of K distinct possible events, such as the various labels that can be assigned to an object during classification. A mass function refers to the level of belief expressed by a source of evidence. This can apply to any subset of the frame of discernment, including the whole frame itself (ignorance state). A particular formalism of the evidence theory by Subjective Logic [17] was used recently as a framework to quantify uncertainty of a neural network [31]. Formally, let K be the number of mutually exclusive singletons with a non-negative belief mass b_K, and overall uncertainty u (belief assigned to the whole frame). More formally:

$$u + \sum_{k=1}^{K} b_k = 1 \tag{4}$$

In other words, b_k is interpreted as the belief mass for the k-th class, whereas u is the total uncertainty mass. Moreover, let $e_k \geq 0$ be the evidence derived for the k-th singleton. The belief b_K and the uncertainty u are computed as:

$$b_k = \frac{e_k}{S} \;\; and \;\; u = \frac{K}{S} \tag{5}$$

where $S = \sum_{k=1}^{K}(e_k + 1)$. In [31], the term *evidence* is a measure from the amount of support collected from data in favor of a sample to be classified into a particular class. Following subjective logic, a belief mass function can be described by a Dirichlet distribution with parameters $\alpha_k = e_k + 1$. In other words, one can derive a belief mass function easily from the parameters of a Dirichlet distribution using $b_k = \frac{(\alpha_k - 1)}{S}$, where $S = \sum_{k=1}^{K}(e_k + 1)$. Hence, the total uncertainty over whole frame u can also be derived.

A typical neural network classifier produces a probability distribution for each sample over the possible classes, using a softmax output layer in most cases. On the other hand, in [31], the authors model the output of a neural network as evidence for a Dirichlet distribution. Let $\boldsymbol{y_i}$ be a one-hot vector encoding the class of observation x_i with $y_{ij} = 1$ and $y_{ik} = 0$ for all $k \neq j$, and p_{ij} is the probability that x_i belongs to the class j, calculated as $p_{ij} = \frac{\alpha_{ij}}{K}$. Finally, the evidential neural network can be trained by minimizing the total MSE loss:

$$L(\Theta) = \sum_{i=1}^{N} \sum_{j=1}^{K} (y_{ij}^2 - 2y_{ij}\,\mathbb{E}[p_{ij}] + \mathbb{E}[p_{ij}^2]) + KL[D(\boldsymbol{p_i}|\boldsymbol{\hat{\alpha}_i})||D(\boldsymbol{p_i}|\langle 1, 1, \ldots, 1\rangle)] \quad (6)$$

where N is the number of training examples. KL is the Kullback-Leibler divergence loss between the Dirichlet distribution of the sample in question with predicted parameters $\hat{\alpha}_i$, and the equivalent of a uniform probability distribution, which is a Dirichlet distribution whose all parameters α_{ij} with $j = 1, 2, \ldots, K$ are equal to 1, and $u = 1$ (total ignorance).

In this work, we adopt the evidential model as a means to acquire valuable information regarding the generated objects of GANs. By incorporating this evidential model, we guide and enhance the training process, enabling us to gain deeper insights into the quality of the generated data.

3 EvGAN: Evidential Generative Adversarial Networks

The architecture of EvGAN, depicted in Figure 1, resembles the original GAN, but with the addition of an auxiliary component. The use of auxiliary information to guide GAN training is a common practice [20,27]. For our method, we employ the evidential neural network (EvNet) [31] described in Section 2.2 as the uncertainty estimator within the GAN architecture. The EvNet is designed to avoid overconfidence in classifying difficult-to-classify objects. Through the KL divergence term in Eq. 6, the evidential model converges to the uniform Dirichlet distribution for misclassified samples. In our case, our goal is to generate objects with high uncertainty that are close to the majority class, that is, a uniform distribution. Therefore, to encourage a conditional GAN to generate samples at the borders of the minority class, we suggest pre-training EvNet on the original data to learn about its distribution. The guided training of GAN is then incorporated by introducing two additional regularization terms, in GAN's loss function. The goal is to ensure the predictive distribution of generated samples has high uncertainty, by acquiring auxiliary knowledge.

3.1 Modified loss function

As discussed previously in Section 2.1, Fisher GAN's objective function was used as base in this paper. The reason behind this choice, is to prevent the issue of mode collapse, as explained previously. The generator's loss function L_G is our only interest in the objective function.

Although the standard GAN model is successful in generating samples from a distribution, it lacks a mechanism to control the specific location of a generated

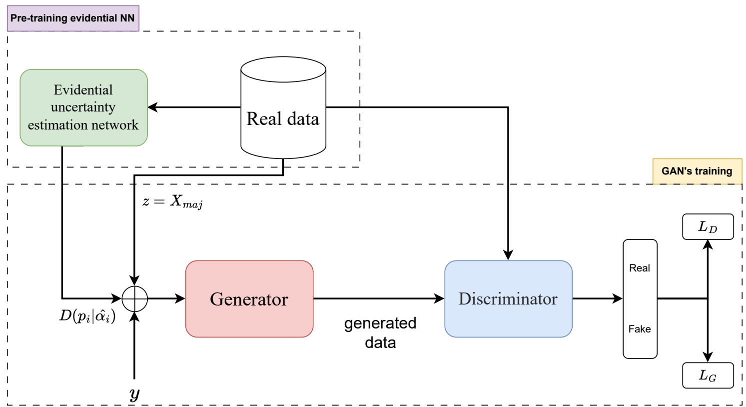


Fig. 1: Overall architecture of EvGAN

output sample based on a given input sample. Consequently, it is not explicitly designed to generate samples with the aim of enhancing imbalanced classification performance.

The introduction of the regularizing loss functions described below allows us to achieve this goal.

KL divergence evidential loss Similarly to the regularization term in EvNet (see Eq. 6), we add a regularization loss function to the generator's loss L_G, called the evidential loss, defined as:

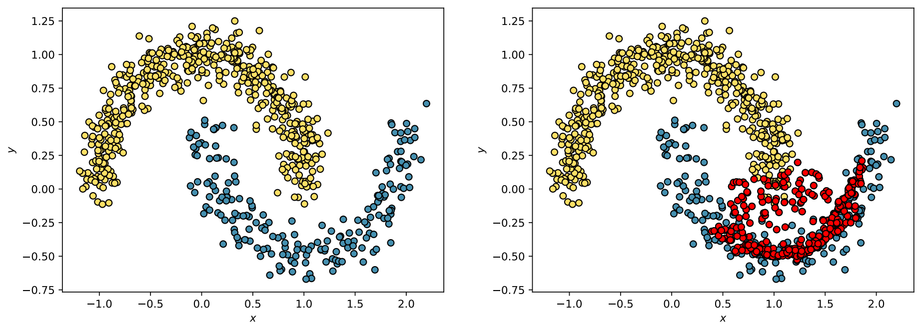
$$\lambda_v \cdot KL[D(\boldsymbol{p_i}|\boldsymbol{\hat{\alpha}_i})||D(\boldsymbol{p_i}|\langle 1, 1, \ldots, 1\rangle)] \tag{7}$$

In this equation, $D(\boldsymbol{p_i}|\boldsymbol{\hat{\alpha}_i})$ represents the Dirichlet distribution predicted by the auxiliary evidential neural network (EvNet) for a generated sample x_i from the generator G. On the other hand, $D(\boldsymbol{p_i}|\langle 1, 1, \ldots, 1\rangle)$ denotes the uniform Dirichlet distribution, where all parameters are equal to 1. The hyperparameter $\lambda_v > 0$ determines the importance of this regularization term.

The purpose of this regularization term is to encourage the GAN to generate samples that are closer to the uniform distribution. By doing so, it promotes the creation of high-uncertainty minority samples. However, it is important to note that this term has the potential to generate noise or outliers that are far from both classes.

To address this concern, we introduce an additional regularization term that mitigates the generation of such undesired samples.

Noise regularization term In many GAN-based approaches, the generator's input is commonly generated from random noise sampled from a latent noise space, often a Gaussian distribution. However, we propose an alternative approach where we feed random real instances from the majority class directly into the



(a) Original distribution (b) After EvGAN oversampling

Fig. 2: A toy imbalanced dataset; yellow points represent the majority class, blue points are the minority one, and the red points are EvGAN-generated.

generator's input. This enables us to incorporate real majority data as additional knowledge within the generator's loss.

To achieve this, we introduce the squared L2 norm to quantify the distance between the input (randomly selected majority data denoted as z) and the generated output (denoted as $G(z)$). This additional term is incorporated into the generator's loss function. Mathematically, the following term is included in the generator's loss:

$$\lambda_z \cdot \left\| z - G(z, y) \right\|^2 \tag{8}$$

where $\lambda_z > 0$ represents a weighting coefficient that determines the significance of the L2 norm term, and y represents the condition label (minority class). By employing this modification, we aim to leverage the information contained in real majority data as an auxiliary component for the generator's training process. The generator will try to minimize this loss, by generating points that are closer to the majority class. Thus, this will complement the evidential term, by ensuring that the highly uncertain objects belong in the space between the majority and minority classes, and not far from both classes, as illustrated in Figure 2.

3.2 Networks' settings

Most research literature on GANs focuses on utilizing image or sequence data, leading to the prevalent use of Convolutional Neural Networks (CNN) [34] or Recurrent Neural Networks (RNN) [23] in the architectures of the generator and discriminator. In contrast, since we focus on tabular datasets in this paper, feed-forward neural networks (FNN) align better with our problem. Therefore, we propose utilizing feed-forward neural networks as the core architecture for our generator, discriminator and evidential models.

Multilayer FNNs are able to learn complex feature interactions. Nevertheless, they might fail to efficiently learn cross feature interactions and discrete features.

Inspired by the work in [25], we propose to set up our networks similarly. Cross layers [36] are added to G, D, and EvNet. This type of layers provides an effective way to model feature interactions by multiplying different input dimensions and learning their relationships. This allows the GAN to capture complex dependencies and correlations among features, improving its ability to generate more diverse and realistic outputs. Through the stacking of N cross layers, we can efficiently calculate feature interactions of up to N degrees in an automated manner. All neural networks are composed of fully connected layers and cross layers.

The generator G employs Leaky ReLU activations for all layers except the last one. The final layer uses a Sigmoid activation for numerical features and Softmax activations for categorical features, with one Softmax activation per feature. Consequently, G is capable of generating distributions for the categorical values of each discrete feature.

Consequently the discriminator D will receive either one-hot encoding vectors from real data, or Softmax distributions for the generated data. Continuous features remain the same for both real and generated data. There is not special processing done for continuous data. However, the distributions of categorical features are transformed into compact, lower-dimensional representations using embeddings. D also uses Leaky Relu activations in all but the last layer, which consists of a dense layer with a sigmoid function.

The network structure of the EvNet model is the same as that of D, except for the final layer. Instead of the original configuration, the EvNet's last layer includes a softmax layer with two outputs, corresponding to the parameters of the evidence which will be used to create Dirichlet distribution's parameters.

4 Experimental study

Having presented our proposed methodology in the preceding section, we now proceed to empirically assess its effectiveness on real-world datasets in this section. Additionally, we compare its performance with that of other baseline methods.

4.1 Experimental setup

Datasets In order to demonstrate the effectiveness by our approach, we conduct experiments on 5 binary real-world datasets from UCI[3] [1] and Kaggle[4]: Online Shoppers Purchasing Intention (*shopping*), *Adult*, Bank Marketing (*bank*), *Coil2000*, and the data mining competition *pakdd2010*. The details of each dataset are summarized in Table 1, where we describe the number of samples, the number of features for each type, and the class distributions. All five tabular datasets have a binary target variable, for which we use the rest of the variables to perform classification. All of the datasets consist of columns that include both numerical

[3] http://archive.ics.uci.edu/ml/datasets

[4] https://www.kaggle.com/competitions/pakdd2010-dataset

and categorical data, underscoring the significance of explicitly considering categorical variables in our approach. To handle missing values, we substitute them with the most commonly occurring value in categorical columns, while numerical columns are assigned the average value of the respective feature.

Table 1: Characteristics of Datasets

Dataset	**#Instances**	**Categorical features**	**Numerical features**	**Class Distribution**
coil2000	9000	25	60	15.76
shopping	12330	8	10	5.5
adult	32561	9	5	3.15
bank	45211	9	6	7.55
pakdd10	46223	27	9	3

Evaluation procedure and metrics To address the inherent imbalance in the benchmarking datasets, we employ a stratified 10-fold cross-validation approach in our evaluation process. We specifically choose a 10-fold setup because GAN-based oversampling techniques often exhibit steep learning curves and require large training sets. In each fold of the cross-validation, we apply oversampling to achieve a balanced parity with a 50:50 ratio. Consistent partitioning of the data is maintained across all oversampling methods to ensure equal difficulty comparisons. To ensure consistency across all methods, we apply min-max scaling to normalize the numerical features within the range of [0, 1]. On the other hand, for handling categorical features, we employ a straightforward approach of one-hot encoding.

Subsequently, we employ a Random Forest classifier [5] to train the model using the resampled dataset. Predictions are then generated using the remaining 10% of the data. To evaluate the performance of each method, we rely on two widely used metrics for imbalanced classification: the Area Under the ROC curve (AUC) [4] score and the Geometric Mean (G-Mean) [3]. These measures provide valuable insights into the effectiveness of the methods in addressing the challenges posed by imbalanced datasets.

Compared methods and Parameters In addition to baseline (no resampling), three benchmark approaches were used for the experiments: SMOTE [6], Borderline SMOTE (B-SMOTE) [15], and the Conditional vanilla GAN (cGAN) [24]. The network configurations for EvGAN are provided in Section 3.2. When using the conditional GAN (cGAN), we adopt similar network settings to our approach, with the exception of excluding cross layers and embeddings in the discriminator. During the training process, we employ the vanilla GAN loss for cGAN.

To find the best regularization coefficients (λ_v and λ_z) for our approach, we perform hyper-parameter tuning using the grid search methodology. This is done on a small validation set for each dataset, allowing us to determine the optimal values. The training process for EvGAN and cGAN adheres to the conventional procedure, with employing the Adam optimization method [18], a commonly used algorithm, with a fixed learning rate of 10^{-4}. As for the other methods we compare against, such as SMOTE and B-SMOTE, we utilize their default parameters.

Implementation The code of our proposal, written in Python 3.9, can be openly found on Github[5]. The implementation uses PyTorch [29] version 2.0.1. The Imbalanced-learn package [21] version 0.11 was used for implementations of benchmark oversampling algorithms and the Scikit-learn package [30] version 1.2.2 was used for supervised learning algorithms and metrics.

4.2 Results discussion

In this section, we present a comprehensive analysis of our method's performance in comparison to other algorithms. The results are displayed in Table 2, showcasing the average G-Mean and AUC scores obtained using a 10-fold stratified cross-validation approach. The best average score is highlighted in bold for easy identification. Notably, when using the Random Forest classifier, our EvGAN method outperformed other algorithms in 4 out of 5 datasets, demonstrating superior performance in both the G-Mean metric and the AUC.

Furthermore, the results highlight the effectiveness of our method in datasets with a large number of categorical features. For datasets like *shopping*, *adult*, and *pakdd10*, regardless of the metric used, EvGAN consistently delivered the best performance. These datasets have more than 10k instances, with *pakdd10* being the largest dataset with 27 categorical features. This demonstrates our method's ability to handle complex datasets and effectively capture relationships between features through our architecture.

The selected metrics, G-Mean and AUC, consider the accuracy of both classes. G-Mean takes into account the true negative rate (specificity) and the true positive rate (sensitivity), while AUC provides a comprehensive measure of overall performance. Thus, we can confidently state that our EvGAN method improves learning on the minority class while maintaining accuracy for the majority class.

5 Conclusion

In this paper, we introduce an innovative oversampling method called evidential GAN, which focuses on strengthening the boundaries of the minority class by generating boundary samples. We leverage the theory of Evidence to guide the training of our GAN, simulating the distribution of minority objects near the boundaries by adding two regularization terms to the generator's loss function. Our approach involves modifying the GAN architecture by incorporating an

[5] https://github.com/faresGr/code-evidential-gan

Table 2: AUC and G-Mean results for chosen datasets using the random forest classifier

	AUC					*G-Mean*				
Datasets	None	SMOTE	B-SMOTE	cGAN	EvGAN	None	SMOTE	B-SMOTE	cGAN	EvGAN
shopping	0.756	0.809	0.814	0.759	**0.836**	0.726	0.800	0.806	0.729	**0.814**
bank	0.691	0.709	0.705	0.716	**0.724**	0.631	**0.692**	0.656	0.635	0.679
adult	0.769	0.769	**0.772**	0.769	**0.772**	0.757	0.769	0.752	0.752	**0.780**
coil2000	0.525	0.536	**0.543**	0.525	0.532	0.258	0.314	0.335	0.258	**0.453**
pakdd10	0.514	0.520	0.518	0.513	**0.537**	0.216	0.262	0.255	0.214	**0.314**

auxiliary evidential component to incorporate uncertainty information into the generator. Additionally, we implement a mechanism to effectively handle tabular data with both numerical and categorical features. The proposed method aims to improve the robustness and performance of GAN-based oversampling for imbalanced datasets.

Finally, the research conducted on benchmark datasets confirmed the effectiveness of the proposed solution. Our experimental study demonstrates that integrating uncertainty quantification by evidence theory into, could result in better robustness of the minority class, which improves the learning performance. Further investigations can include applying our framework to generate minority class data in more complex distributions such unstructured data, i.e., images and time series.

References

1. Asuncion, A., Newman, D.: UCI machine learning repository (2007)
2. Aung, M.H., Seluka, P.T., Fuata, J.T.R., Tikoisuva, M.J., Cabealawa, M.S., Nand, R.: Random forest classifier for detecting credit card fraud based on performance metrics. In: 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE). pp. 1–6 (2020). https://doi.org/10.1109/CSDE50874.2020.9411563
3. Barandela, R., Valdovinos, R.M., Sánchez, J.S.: New applications of ensembles of classifiers. Pattern Analysis & Applications **6**(3), 245–256 (2003)
4. Bradley, A.P.: The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern Recognition **30**(7), 1145 – 1159 (1997)
5. Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)
6. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research **16**, 321–357 (2002)
7. Chen, J., Pi, D., Wu, Z., Zhao, X., Pan, Y., Zhang, Q.: Imbalanced satellite telemetry data anomaly detection model based on bayesian LSTM. Acta Astronautica **180**, 232–242 (2021)
8. Cover, T., Hart, P.: Nearest neighbor pattern classification. IEEE transactions on information theory **13**(1), 21–27 (1967)
9. Cui, J., Zong, L., Xie, J., Tang, M.: A novel multi-module integrated intrusion detection system for high-dimensional imbalanced data. Applied Intelligence **53**(1), 272–288 (2023)

10. Dempster, A.P.: A generalization of bayesian inference. Journal of the Royal Statistical Society: Series B (Methodological) **30**(2), 205–232 (1968)
11. Engelmann, J., Lessmann, S.: Conditional Wasserstein GAN-based oversampling of tabular data for imbalanced learning. Expert Systems with Applications **174**, 114582 (2021)
12. Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B., Herrera, F.: Learning from imbalanced data sets, vol. 10. Springer (2018)
13. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K. (eds.) Advances in Neural Information Processing Systems. vol. 27, pp. 2672–2680 (2014)
14. Grina, F., Elouedi, Z., Lefevre, E.: Re-sampling of multi-class imbalanced data using belief function theory and ensemble learning. International Journal of Approximate Reasoning **156**, 1–15 (2023)
15. Han, H., Wang, W.Y., Mao, B.H.: Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. Lecture Notes in Computer Science **3644**(PART I), 878–887 (2005)
16. He, H., Bai, Y., Garcia, E.A., Li, S.: ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence). pp. 1322–1328. IEEE (2008)
17. Jøsang, A.: Subjective Logic: A Formalism for Reasoning Under Uncertainty. Artificial Intelligence: Foundations, Theory, and Algorithms, Springer Publishing Company, Incorporated, 1st edn. (2016)
18. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: ICLR (2015)
19. Koziarski, M., Bellinger, C., Woźniak, M.: RB-CCR: Radial-based combined cleaning and resampling algorithm for imbalanced data classification. Machine Learning **110**, 3059–3093 (2021)
20. Lee, K., Lee, H., Lee, K., Shin, J.: Training confidence-calibrated classifiers for detecting out-of-distribution samples. arXiv preprint arXiv:1711.09325 (2017)
21. Lemaître, G., Nogueira, F., Aridas, C.K.: Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. The Journal of Machine Learning Research **18**(1), 559–563 (2017)
22. Li, D., Zheng, C., Zhao, J., Liu, Y.: Diagnosis of heart failure from imbalance datasets using multi-level classification. Biomedical Signal Processing and Control **81**, 104538 (2023)
23. Li, X., Metsis, V., Wang, H., Ngu, A.H.H.: TTS-GAN: A transformer-based time-series generative adversarial network. In: 20th International Conference on Artificial Intelligence in Medicine, AIME 2022, Halifax, NS, Canada, June 14–17, 2022, Proceedings. pp. 133–143. Springer (2022)
24. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
25. Mottini, A., Lheritier, A., Acuna-Agost, R.: Airline passenger name record generation using generative adversarial networks. arXiv preprint arXiv:1807.06657 (2018)
26. Mroueh, Y., Sercu, T.: Fisher GAN. Advances in neural information processing systems **30**, 2513–2523 (2017)
27. Odena, A., Olah, C., Shlens, J.: Conditional image synthesis with auxiliary classifier gans. In: International conference on machine learning. pp. 2642–2651. PMLR (2017)

28. Park, N., Mohammadi, M., Gorde, K., Jajodia, S., Park, H., Kim, Y.: Data synthesis based on generative adversarial networks. arXiv preprint arXiv:1806.03384 (2018)
29. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems 32, pp. 8024–8035. Curran Associates, Inc. (2019), http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf
30. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
31. Sensoy, M., Kaplan, L., Kandemir, M.: Evidential deep learning to quantify classification uncertainty. Advances in neural information processing systems **31** (2018)
32. Shafer, G.: A mathematical theory of evidence, vol. 42. Princeton university press (1976)
33. Smets, P.: The Transferable Belief Model for Quantified Belief Representation, pp. 267–301. Springer Netherlands, Dordrecht (1998)
34. Torbunov, D., Huang, Y., Yu, H., Huang, J., Yoo, S., Lin, M., Viren, B., Ren, Y.: Uvcgan: Unet vision transformer cycle-consistent GAN for unpaired image-to-image translation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 702–712 (2023)
35. Vuttipittayamongkol, P., Elyan, E., Petrovski, A.: On the class overlap problem in imbalanced data classification. Knowledge-based systems **212** (2020)
36. Wang, R., Fu, B., Fu, G., Wang, M.: Deep & cross network for ad click predictions. In: Proceedings of the ADKDD'17, pp. 1–7 (2017)